

## **MACHINE LEARNING**

1. b) 4
2. d) 1, 2 and 4
3. d) formulating the clustering problem
4. a) Euclidean distance
5. b) Divisive clustering
6. d) All answers are correct
7. a) Divide the data points into groups
8. b) Unsupervised learning
9. d) All of the above
10. a) K-means clustering algorithm
11. d) All of the above
12. a) Labeled data

### **13. How is cluster analysis calculated?**

K means Clustering follows the below steps

- First, an initial partition with k clusters (given number of clusters) is created.
- Then, starting with the first object in the first cluster, Euclidean distances of all objects to all cluster foci are calculated.
- If an object is detected whose distance to the center of gravity of the own cluster is greater than the distance to the center of gravity (centroid) of another cluster, this object is shifted to the other cluster.
- Finally, the centroids of the two changed clusters are calculated again, since the compositions have changed here.
- These steps are repeated until each object is located in a cluster with the smallest distance to its centroid (center of the cluster) (optimal solution).

The hierarchical cluster analysis follows three basic steps

- Calculating the distances
- Linking the clusters
- Choosing a solution by selecting the right number of clusters.

### **14. How is cluster quality measured?**

The cluster quality is generally measured using average silhouette coefficient value of all objects in the data set. It will evaluate the goodness of clustering by considering how well the clusters are separated and how compact the clusters are.

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where  $a(i)$  is the average distance of point  $i$  from all other points in its cluster and  $b(i)$  is the smallest average distance of  $i$  to all points in any other cluster.

### **15. What is cluster analysis and its types?**

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same cluster are more similar to each other than to those in other clusters.

Types of clustering :

- Centroid-based clustering
- Density-based clustering
- Distribution-based clustering
- Connectivity-based clustering
- Grid-based clustering